Submit on/before 25/07/2023

Summary Measures
Assignment 02
UWU/CST/21/038

1. For these situations, state which measure of central tendency should be used.
   a. The most typical case is desired - **Mode**
   b. The distribution is open-ended - **Median**
   c. There is an extreme value in the set - **Median**
   d. The data are categorical - **Mode**
   e. Further statistical computations will be needed - **Mean**

2. The following data are the heights, correct to the nearest centimeter, for a group of children.

144	132	138	129	135	137	143	152	126	137
161	133	129	132	133	146	141	154	147	136

   a. Draw a stem and leaf diagram of the data and discuss the skewness of the distribution

Stem	Leaf
12	6,9,9
13	2,2,3,3,5,6,7,7,8
14	1,3,4,6,7
15	2,4
16	1

Mode = 137

Median = 137

Mean = 139.25

(Mode=Median)<Mean

So this is **Positive** Skewed.

Ascending Order:-
126,129,129,132,132,133,133,135,136,137,137,138,141,143,144,146,147,152,154,161

   b. Find the median and the inter quartile range of the data.

Median=((n+1)/2)th
= ((20+1)/2)th
=10.5th place
=137+(137-137) *0.5 =$\underline{\underline{137}}$

Inter quartile-Range=Q3-Q1

Q3=(((n+1)/4) *3)th
= (((20+1)/4) *3) th
=15.75 th
=144+(146-144) *0.75
=144+1.5

=145.5

Q1=((n+1)/4) th Value

= ((20+1)/4)thValue

=5.25 th Value

=132+(133-132) *0.25

=132+0.25

=132.25

Inter quartile-Range=Q3-Q1

=145.5-132.25

=13.25

c. Calculate the average height of children

Average height=

(126+129+129+132+132+133+133+135+136+137+137+138+141+143+144+146 +147+152+154+161)/20

=2785/20

=139.25

d. Which is the most suitable location parameter to describe the data?

Q2(Median)-The most suitable parameter to describe the data is Q3 that mean median. The median is not affected by extreme values.

e. Calculate the suitable dispersion measurement.

= (20+1)/2

=0.5

=137+(137-137) *0.5

=137

3. Calculate the mean, mode, median, first quartile & third quartile and Standard Deviation for the following distribution,

Class Interval	10—19	20—29	30—39	40—49	50—59	60—69	70—79	80—89
Frequency	2	4	9	11	12	6	4	2

Class Interval(X)	**Frequency(F)**	**Mid Value(Xi)**	**FXi**	**Cumulative Frequency**
10-19	2	14.5	29	2
20-29	4	24.5	98	6
30-39	9	34.5	310.5	15
40-49	11	44.5	489.5	26
50-59	12	54.5	654	38
60-69	6	64.5	387	44
70-79	4	74.5	298	48
80-89	2	84.5	169	50
	50		2435	

Mean=2435/50
= 48.7

Because 50/2,25 is near to cumulative frequency 26, which belongs to class interval 40–49, the median falls inside this range.

Median=L+((N/2-C)/F) *h
=40+((25-15)/11) *9
=48.18

Mode=L+((F1-F0)/(2F1-F0-F2)) *h
=50+((12-11)/ (2*12-11-6)) *9
=51.3

Q1=30+(12.5-6)/9*9 =36.5

Q3=50+(37.5-26)/12*9 =58.63

Standard Deviation=16.79

4. A seed merchant took delivery of a sack of beans. Looking into the sack, he thought the beans appeared much more variable in size than his usual stock. He took a sample of 15 beans from the sack and measured their lengths. His results were as follows, in cm.
Calculate the Range, Mean, Standard deviation and Coefficient of Variation for these 15 beans and interpret the results.

2.4, 2.7, 3.0, 2.5, 3.2, 2.8, 2.8, 2.3, 3.0, 2.8, 2.8, 2.4, 2.9, 2.8, 3.1

2.3,2.4,2.4,2.5,2.7,2.8,2.8,2.8,2.8,2.8,3.0,3.0,3.1,3.2

Range = 3.2-2.3=0.9
Mean= (2.3+2.4+2.4+2.5+2.7+2.8+2.8+2.8+2.8+2.8+2.9+3.0+3.0+3.1+3.2)/15 =2.77
Standard Deviation=0.2664
Coefficient of Variation= (0.2664/2.7) *100% =9.6278%

5. The boxplots (Figure 01) shown below represent the men and women age distributions of Olympic Soccer Team Players in 2016. What can you say about the below measurements of the distributions given the boxplots above?

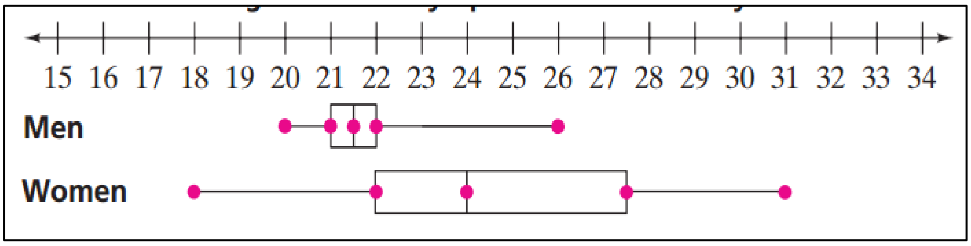


Figure 1: Ages of Olympic Soccer Team Players

a. Location

Men's age distribution is centered between 21 and 22 whereas women's age distribution is centered at around 24.

b. Shape

Median Q2 is seem so close to the first quartile Q1 and third quartile Q3 with a small difference men’s age distribution seems to be relatively symmetrical. In the women’s age distribution third quartile Q3 is larger than median Q2.Therefore women’s age distribution seems to be positively skewed.

c. Spread

The women’s age distribution has a larger spread comparing to the men’s age distribution

d. Kurtosis

6. The index number of prices of cotton and coal shares in 2013 is given below:

Month	January	February	March	April	May	June	July	August	Sep.	Oct.	Nov.	Dec.
cotton shares	188	178	173	164	172	183	184	185	211	217	232	240
coal shares	131	130	130	129	129	120	127	127	130	137	140	142

Which Share index is more variable?

Range for cotton shares=240-164=76
Range for coal shares =142-120=22
Standard Deviation for cotton shares=28.32
Standard Deviation for coal shares=4.25

index of cotton share is more variable than the index of coal shares.

7. From the marks secured by 120 students in class A and class 13, the following measures are obtained:

Class	Mean	Standard deviation	Mode
A	46.83	14.8	51.67
B	47.83	14.8	47.07

Determine which distribution of marks is more skewed.

Skewness=3*(Mean-Mode)/Standard Deviation

Skewness for class A=3*(46.83-51.67)/14.8 =-0.9810
Skewness for class B =3*(47.83-47.07)/14.8 =0.154

Now we can compare the absolute values of both class A and class B,
Class A=|-0.981|=0.981

Class B=|0.154|=0.154

class A is more skewed than the distribution of marks in class B.

8. The information below is derived from the National Diet and Nutrition Survey of British Adults, 2008–2009. It relates to a random sample of adults aged 19–64 in private households in mainland Britain.

Age (years)	**Number of Men**	**Number of Women**
19–24	61	78
25–34	160	211
35–54	393	486
55–64	152	183
Total	766	958
Region		
Scotland and the North	248	326
Central, South West and Wales	274	350
London and the South East	244	282
Current smoker	236	313

Write a report comparing the representation of men and women within the various categories in this achieved sample. Your report should include diagrams and summary statistics. You should hand in any calculations needed to produce the diagrams and summary statistics, making clear that these calculations are not part of the report.

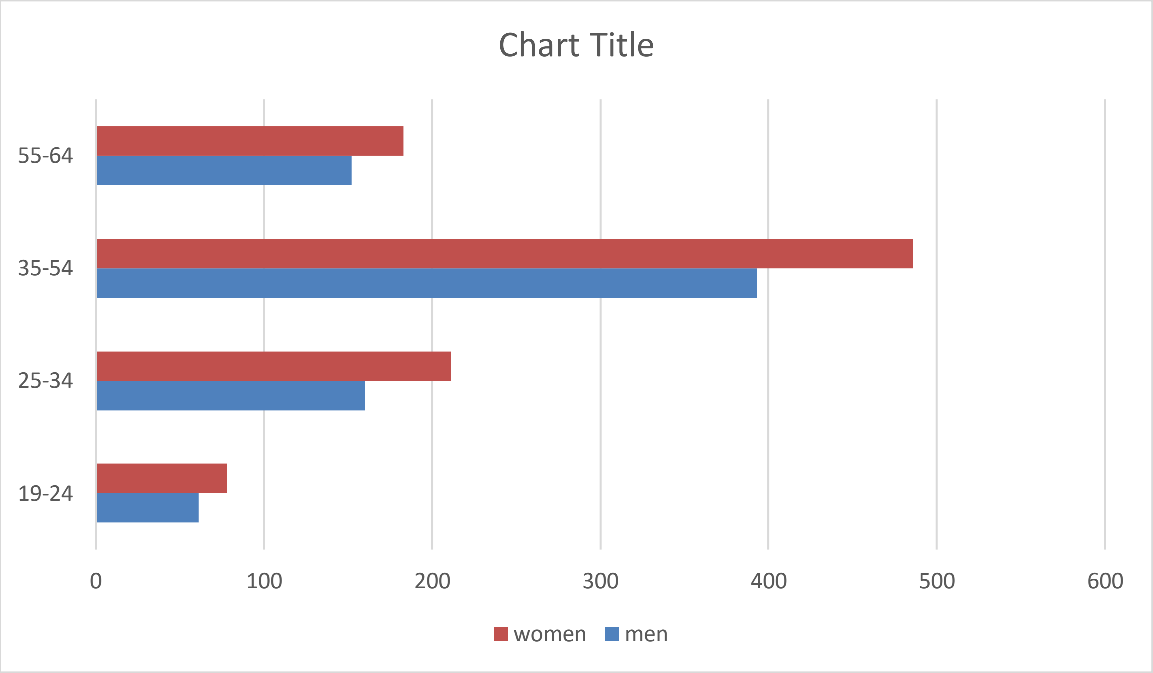
Introduction:

The National Diet and Nutrition Survey of British Adults in 2008-2009 aimed to understand the dietary. Habits and nutrition of adults aged 19-64 in private households in mainland Britain. This report comparing the representation of men and women in various age groups and regions, as well as their smoking status.

Data:

The data was collected from a random sample of adults, with the total number of participants being 766 men and 958 women.

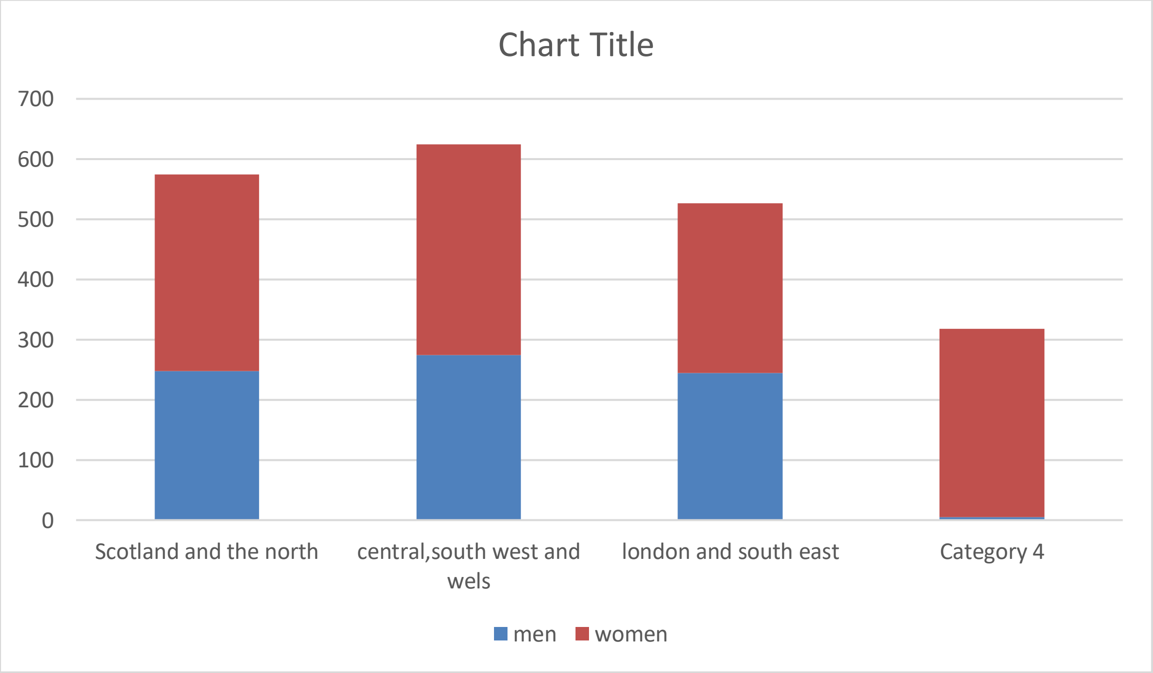
Representation by age group:



From the above bar chart, we can say that:

- Among the participants aged 19-24, there were 61 men and 78 women.
- In the 25-34 age group, there were 160 men and 211 women.
- The 35-54 age group had the largest representation, with 393 men and 486 women.
- In the 55-64 age group, there were 152 men and 183 women

Representation by Region:

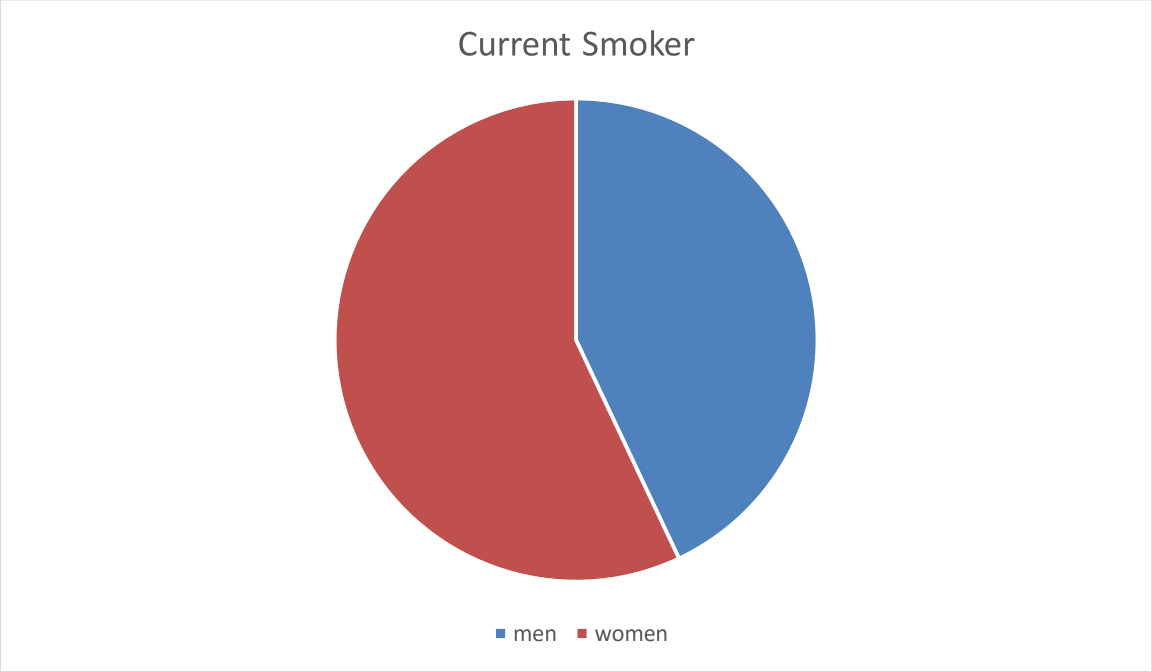


From the above stacked bar chart, we can say that:

- In the Scotland and North region, there were 248 men and 326 women.

- The central, south west and Wales region had 274 men and 350 women.
- In the London and South East region, there were 244 men and 282 women

Representation of Smoking:



From the above Pie Chart, we can observe that:

- Out of the 766 men,236 were current smokers.
- Among the 958 women, 313 were current smokers.

Summary Conclusion:

- Age group 19-24: Men-61, Women-78
- Age group 25-34: Men-160, Women-211
- Age group 35-54: Men-393, Women-486
- Age group 55-64: Men-152, Women-183
- Region Scotland and the North: Men-248, Women-326
- Region central, south west and Wales: Men-274, Women-350
- Region London and the south east: Men-244, Women-282
- Current Smokers: Men-236, Women-313

Conclusion

- We can observe that men and women are well represented in all age groups and regions, with no significant disparities. And the data indicates that there are more current male smokers than female smokers in the sample.

9. Suppose that the following two data sets I and II correspond to percentage marks obtained by two groups of students of size twenty each. They are arranged in ascending order of magnitude.

Student no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

I	11	11	13	14	18	21	23	28	30	33	37	39	41	43	54	57	79	81	82	87
II	10	22	27	28	28	29	29	29	31	35	38	38	50	52	54	80	82	90	90	95

a. Classify each of the above data sets according to the following table.

Class Boundaries	Frequency I	Frequency II
E: (0.05- 29.5)	8	8
D: (29.5- 39.5)	4	4
C: (39.5- 54.5)	3	3
B: (54.5- 69.5)	1	0
A: (69.5- 99.5)	4	5

b. Hence calculate the means and standard deviation using (i) raw data and (ii) classified data for each of the data sets. Compare the accuracies and comment critically on the assumption about

Raw data:

For Data Set I:

Mean I = (11 + 11 + 13 + 14 + 18 + 21 + 23 + 28 + 30 + 33 + 37 + 39 + 41 + 43 + 54 + 57 + 79 + 81 + 82 + 87) / 20

= 802 / 20 ≈ 40.1

For Data Set II:

Mean II = (10 + 22 + 27 + 28 + 28 + 29 + 29 + 29 + 31 + 35 + 38 + 38 + 50 + 52 + 54 + 80 + 82 + 90 + 90 + 95) / 20

= 937 / 20 ≈ 46.85

Classified data:

For Data Set I:

Mean I = [(8 * 14.75) + (4* 34.5) + (3* 47) + (1 * 62) + (4* 84.5)] / 20

= (118+ 138 + 141+ 62+ 338) / 20 ≈ 39.85

For Data Set II:

Mean II = [(8 * 14.75) + (4 * 34.5) + (3 * 47) + (0* 62) + (5* 84.5)] / 20

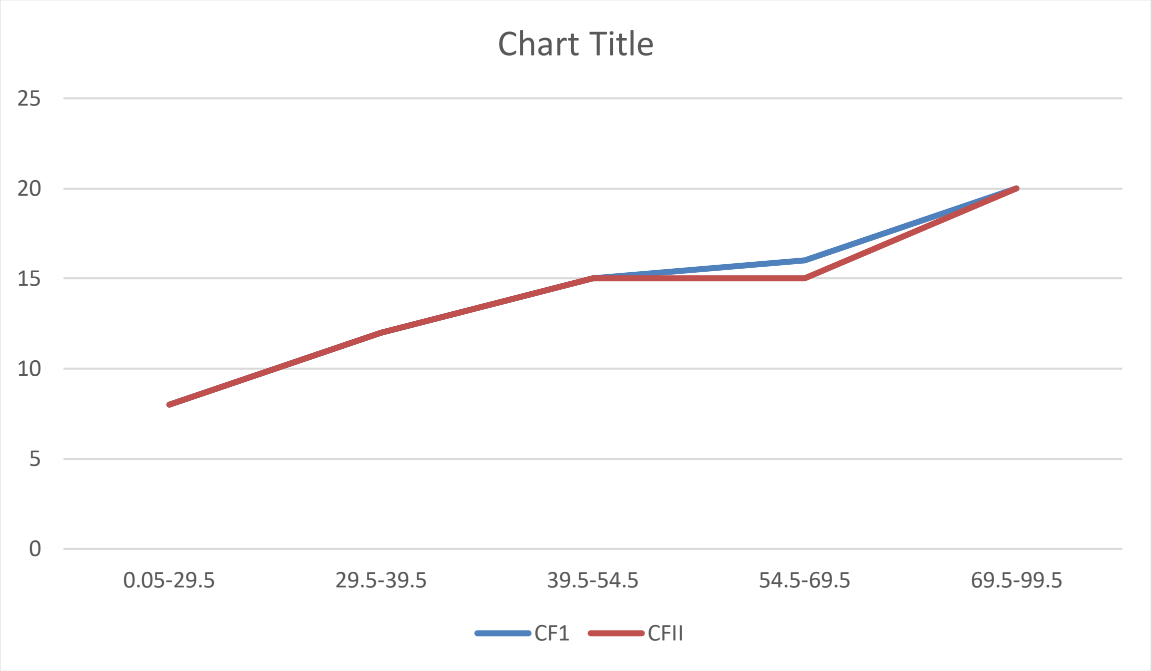
= (118+138+141+0+422.5) / 20 ≈ 40.97

c. Find the cumulative frequencies for each group of students and draw their graphs on the same graph paper. Comment on the differences.

Class Boundaries	E: (0.05-29.5)	D: (29.5-39.5)	C: (39.5-54.5)	B: (54.5-69.5)	A: (69.5-99.5)
Frequency I	8	4	3	1	4
Cumulative Frequency I	8	12	15	16	20

Class Boundaries	E: (0.05-29.5)	D: (29.5-39.5)	C: (39.5-54.5)	B: (54.5-69.5)	A: (69.5-99.5)
Frequency II	8	4	3	0	5

Cumulative Frequency II	8	12	15	15	20



01. Due to the shortage of coins below the five rupee denomination and the inconvenience arising from it, it has been proposed by a statistician to charge Rs. 10 for both “Rs. 8” and “Rs. 11” distances and Rs. 15 for all the distance charged at Rs. 13,15 and 17 at present: no change being made for Rs. 5 tickets. In order to study the feasibility of this proposal, sixty (60) trips each of three passengers A, B and C are considered. The following table gives their old and new travel patterns fares in Rupees.

Bus Fare(Rs.)		**Number of Trips**		
Old	**New**	**Passenger A**	**Passenger B**	**Passenger C**
5	5	5	10	0
8	10	5	15	5
11	10	10	15	15
13	15	15	10	10
15	15	15	5	10
17	15	10	5	20
Total No of Trips		**60**	**60**	**60**

a. What is the net return to the Bus owner from the three passengers due to new scheme?

Total Expenditure under the Old Scheme:

Passenger A:

Total expenditure = 5 * 5 + 8 * 5 + 11 * 10 + 13 * 15 + 15 * 15 + 17 * 10

= 25 + 40 + 110 + 195 + 225 + 170

= Rs.765

Passenger B:

Total expenditure = 5 * 10 + 8 * 15 + 11 * 15 + 13 * 10 + 15 * 5 + 17 * 5

= 50 + 120 + 165 + 130 + 75 + 85

= Rs.625

Passenger C:

Total expenditure = 5 * 0 + 8 * 5 + 11 * 15 + 13 * 10 + 15 * 10 + 17 * 20

= 0 + 40 + 165 + 130 + 150 + 340

=Rs. 825

Total = 765 + 625 + 825

Total = 2215

Total Expenditure under the New Scheme:

Passenger A:

Total expenditure = 5 * 5 + 10 * 5 + 10 * 10 + 15 * 15 + 15 * 15 + 15 * 10

= 25 + 50 + 100 + 225 + 225 + 150

= Rs.775

Passenger B:

Total expenditure = 5 * 10 + 10 * 15 + 10 * 15 + 15 * 10 + 15 * 5 + 15 * 5

= 50 + 150 + 150 + 150 + 75 + 75

= Rs.650

Passenger C:

Total expenditure = 5 * 0 + 10 * 5 + 10 * 15 + 15 * 10 + 15 * 10 + 15 * 20

= 0 + 50 + 150 + 150 + 150 + 300

=Rs. 800

Total = 775 + 650 +800

Total = 2225

Total Expenditure under the Old Scheme: Rs. 2215

Total Expenditure under the Old Scheme: Rs. 2225

Net Return = 2225 - 2215

Net Return = Rs. 10

b. Calculate the means and modes of charges for all three passengers **both schemes** and comment on the skewness in each case using suitable diagram or chart only.

Passenger A:

- Mean = [(5 * 5 + 10 * 5 + 10 * 10 + 15 * 15 + 15 * 15 + 15 * 10) / 60] = 12.91
- Mode = 15

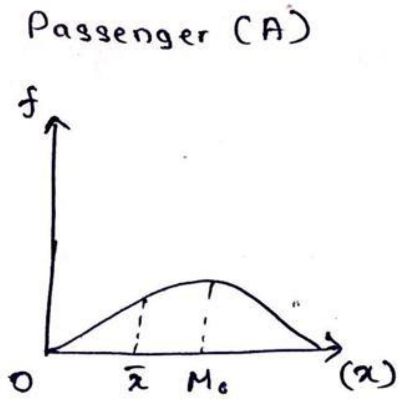
Passenger B:

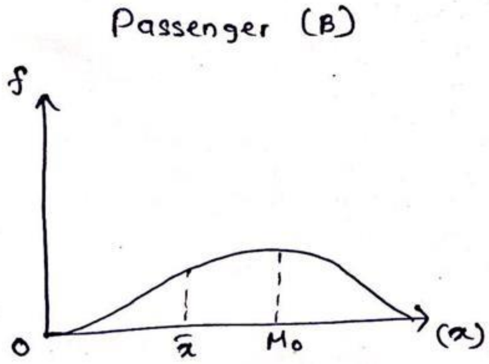
- Mean = [(5 * 10 + 10 * 15 + 10 * 15 + 15 * 10 + 15 * 5 + 15 * 5) / 60] = 10.83
- Mode = 15

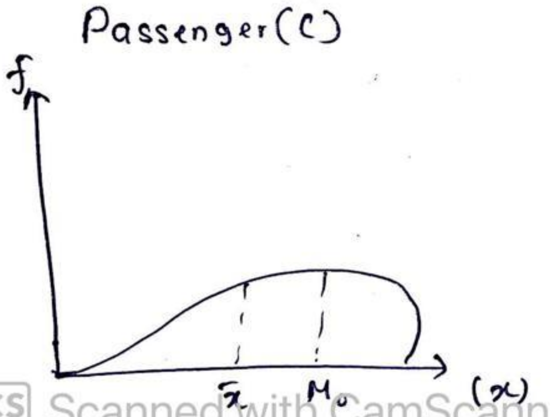
Passenger C:

- Mean = [(5 * 0 + 10 * 5 + 10 * 15 + 15 * 10 + 15 * 10 + 15 * 20) / 60] = 13.33
- Mode = 20

- Passenger A: -Mode>Mean
- Passenger B: -Mode>Mean
- Passenger C: -Mode>Mean







All the Three Graphs show negative skewness. Because of Mode>Mean